

Identification and Estimation Issues in Exponential Smooth Transition Autoregressive Models*

DANIEL BUNCIC

Research and Modelling Division, Financial Stability Department, Sveriges Riksbank, SE-103 37, Stockholm, Sweden (email: daniel.buncic@riksbank.se, http://www.danielbuncic.com)

Abstract

Exponential smooth transition autoregressive (ESTAR) models are widely used in the international finance literature, particularly for the modelling of real exchange rates. We show that the exponential function is ill-suited as a regime weighting function because of two undesirable properties. Firstly, it can be well approximated by a quadratic function in the threshold variable whenever the transition function parameter γ , which governs the shape of the function, is ‘small’. This leads to an identification problem with respect to the transition function parameter and the slope vector, as both enter as a product into the conditional mean of the model. Secondly, the exponential regime weighting function can behave like an indicator function (or dummy variable) for very large values of γ . This has the effect of ‘spuriously overfitting’ a small number of observations around the location parameter μ . We show that both of these effects lead to estimation problems in ESTAR models. We illustrate this by means of an empirical replication of a widely cited study, as well as a simulation exercise.

I. Introduction

The Exponential Smooth Transition Autoregressive (ESTAR) model has become one of the workhorse econometric models in the international finance literature, particularly for the modelling of real exchange rates. ESTAR models were introduced by Granger and Teräsvirta (1993) and Teräsvirta (1994) into the economics literature as a generalization

JEL Classification numbers: C13, C15, C50, F30, F44.

*I am grateful to the editor, two anonymous referees as well as Lorenzo Camponovo, Adrian Pagan, Paolo Giordani, Timo Teräsvirta, Xin Zhang, Dick van Dijk, Paul Karehnke (Discussant), Pedro Barroso, Simon van Norden, Robert Taylor, Jonas Striaukas, Leonardo Iania, Annastiina Silvennoinen, Irina Panovska, Luc Bauwens, Fabio Canova and seminar and conference participants at the Université catholique de Louvain (CORE), the University of St. Gallen, Sveriges Riksbank, the Australasian Finance and Banking Conference in Sydney (2017), the International Conference on Time Series Econometrics in Granada (2017), the 23rd EBES Conference in Madrid (2017), the 26th Annual SNDE Symposium in Tokyo (2018), and the IAAE 2018 Annual Conference in Montreal for their comments that helped to improve the paper. Parts of this paper were written while I was visiting the Research Division of the Monetary Policy Department at the Riksbank. I thank Jesper Lindé and his colleagues for their hospitality and many stimulating discussions. The opinions expressed in this article are the sole responsibility of the author and should not be interpreted as reflecting the views of Sveriges Riksbank.

of the (nonlinear) exponential autoregressive model of Haggan and Ozaki (1981) and threshold time series models of Tong (1983).

Despite being widely used, the exponential function employed in ESTAR models is not suitable as a regime weighting function. The reason for this being two undesirable features of the exponential function. The first is that for small values of the transition function parameter γ , the exponential function can be well approximated by a quadratic in the threshold variable z_t . This leads to the slope vector attached to the nonlinear regime and the γ parameter entering as a product into the nonlinear conditional mean of the model, which leads to identification issues. What is particularly problematic with this scenario is that it is not a small sample issue but rather a property of the model and is evident even in very large samples of 5,000 observations.

The second undesirable feature is that for extremely large values of γ , the exponential weighting function will be equal to unity for nearly all values of the transition variable z_t , except when z_t is equal to the location parameter μ . The effect of this on the model is that only a very small number of observations around μ receive a weight that is not equal to unity, which leads to an ‘*outlier fitting effect*’ of the exponential function. Although this is evident in ‘*small*’ samples only, our simulation results indicate that it can be pervasive for sample sizes as large as 500 observations, resulting in ‘*large*’ γ estimates for over 70% of the simulations.

There exists ample evidence of these problems in empirical studies. For instance, Michael, Nobay and Peel (1997) fit ESTAR models to real exchange rate data for a number of countries. In panels (a) and (b) of Figure 1 on page 875 in their paper, one can see that for the UK–US series, the weighting function remains well below 0.3 for the entire range of the data, while for the UK–France series, only 4 data points receive a weight in excess of 0.3, with both functions being quadratic looking in shape. Taylor and Peel (2000) use monetary fundamentals to study the evolution of exchange rates and utilize the ESTAR model to capture nonlinearities in the data. From the regime weighting functions plotted in Figure 2 on page 45 of their paper, it can be seen that the transition function weights remain below 0.4 over the entire range of the data and are again quadratic looking. The study by Baum, Barkoulas and Caglayan (2001) provides even stronger symptoms of a weakly identified model. The estimates of the transition function parameter γ that are reported in Tables 4 and 5 on page 391 of Baum *et al.* (2001) are — with the exception of the WPI-based real exchange rate for Norway — between 0.0042 and 0.0833! The corresponding transition function plots, on pages 392 and 393, show a quadratic shape.

Similar issues are evident in Sarantis (1999), Taylor, Peel and Sarno (2001), Kilian and Taylor (2003), Kapetanios, Shin and Snell (2003), Sarno, Leon and Valente (2006), Paya and Peel (2006), Sollis (2008), Taylor and Kim (2009), Cerrato, Kim and Macdonald (2010), Pavlidis, Paya and Peel (2011), Beckmann, Berger and Czudaj (2015) and many others. The study by Beckmann *et al.* (2015) is particularly noteworthy to single out here, as the estimation results reported in Table 3 of their paper provide first hand empirical evidence of both estimation problems that we outline above. Beckmann *et al.* (2015) estimate ESTAR models on gold returns, using stock returns from 23 different equity markets as regressor and threshold variables. The model is complicated by the addition of a GARCH type volatility process on the error term in the ESTAR models that are fitted. From Table 3 on page 22 in their paper, we can see that the estimates of the transition function parameter γ

hit the lower bound of 0 for 5 out of the 23 results, with the corresponding slope parameters becoming extremely large in absolute value. For 3 out of 23, the γ estimates are very large, with the one for the World equity market being over 27,000! These findings are clear symptoms of the type of identification and estimation issues that make ESTAR models unsuitable for econometric modelling.

The objective of this study is to provide a detailed analysis of these identification and estimation issues with ESTAR models. We begin by showing analytically through a Taylor series expansion of the exponential function that if the higher order terms in the auxiliary model are zero, the model is not identified with respect to the slope vector and the transition function parameter. We illustrate the identification and estimation problems in an empirical setting by replicating the well known and widely cited study of Taylor *et al.* (2001). Lastly, we present a simulation analysis, where we assess the severity of the above discussed issues with regards to increasing values in the γ parameter and sample sizes in a controlled simulation experiment.

The remainder of the paper is organised as follows. Section II provides a brief description of the ESTAR model. In section III, we formalise the identification issue and outline the tools that we use to measure weak identification. In section IV we illustrate these problems by replicating a well known empirical study on nonlinear mean reversion in exchange rates. In section IV we also perform a simulation study where we vary the size of the γ parameter as well as the sample size. We conclude the study in section V, where we also offer suggestions to practitioners what alternative weighting functions could be used instead of the exponential function. In an Online Appendix we provide additional material that may be of interest to the applied researcher and practitioner.

II. ESTAR models

Let y_t be a scalar time series which follows a general Smooth Transition Autoregressive (STAR) model, taking the form:¹

$$y_t = \mathbf{x}_t \boldsymbol{\alpha} + \mathbf{x}_t \boldsymbol{\beta} \mathcal{G}(z_t; \gamma, \mu) + \epsilon_t, \tag{1}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(k \times 1)$ vectors of regime specific slope parameters and ϵ_t is a disturbance term assumed to follow a martingale difference sequence (MDS) with respect to the information set \mathcal{F}_{t-1} , with mean 0 and variance σ^2 . The $(1 \times k)$ vector of control (or regressor) variables at time t is denoted by \mathbf{x}_t .² The regime weighting function $\mathcal{G}(z_t; \gamma, \mu)$ is a continuous and smooth function, which is bounded between $[0,1]$. When it takes the

¹ It is common to interpret STAR type models either within a regime switching framework, where the transition from one regime to the other is smooth, or they can be viewed as a continuum of regimes, yielding a general parametric nonlinear conditional mean function. In this paper, we will stick to the regime narrative as it fits well with the economic motivation and data that STAR models have been fitted to (see also van Dijk, Teräsvirta and Franses (2002) for a discussion of this view and Buncic (2012) for an out-of-sample forecast evaluation. Other models that have been fitted to exchange rate data try to model the nonlinearity with time-varying parameter models (see for instance Buncic and Piras (2016)).

² Note here that, in order to simplify the description of the model, we do not use different notation for the regressors \mathbf{x}_t in the two ‘regimes’. That is, we do not write $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ to emphasize that they could be different. It is implicitly assumed that these regressors can be a subset of the global regressor set $\mathbf{x}_t = [1 \ y_{t-1} \ \mathbf{w}_{t-1} \ \mathbf{d}_t]$, where \mathbf{y}_{t-1} is a vector of lagged dependent variables, \mathbf{w}_{t-1} is a vector of lagged exogenous variables, and \mathbf{d}_t is a vector of predetermined or deterministic time trend polynomials, dummy variables and/or seasonal indicators known for all t . In empirical

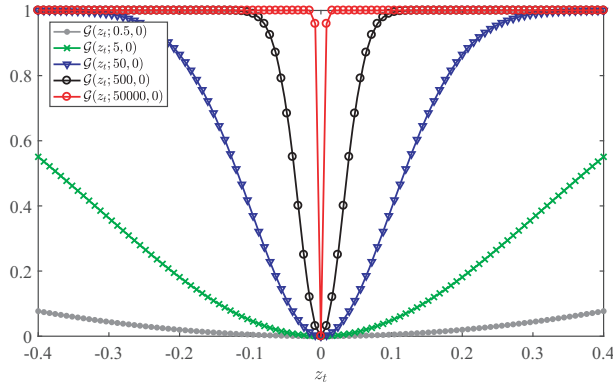


Figure 1. Exponential transition weighting function $\mathcal{G}(z_t; \gamma, \mu)$ plotted over 101 equally spaced z_t values

form of an exponential function defined as $\mathcal{G}(z_t; \gamma, \mu) = 1 - \exp\{-\gamma(z_t - \mu)^2\}$, the model in (1) is known as the Exponential STAR (ESTAR) model. The parameter $\gamma \in \mathbb{R}_+$ in $\mathcal{G}(z_t; \gamma, \mu)$ determines the smoothness of the transition function, while $\mu \in \mathbb{R}$ is a threshold location parameter. The transition variable z_t can be a deterministic variable, an exogenous variable known at time $t - 1$, or, as is quite frequently the case in empirical studies, the lagged endogenous variable, that is, $z_t = y_{t-q}$, for some integer $q > 0$.

The exponential function is plotted in Figure 1 below. We have normalized μ to 0, and show the function over a commonly encountered interval from -0.4 to 0.4 for threshold variable z_t and an equally spaced grid of 101 points, for five different γ values ($\gamma = \{0.5, 5, 50, 500, 50,000\}$). The limiting properties of the exponential weighting function are visible graphically. For instance, one can notice that for the two ‘smaller’ values of γ of 0.5 and 5, the exponential function $\mathcal{G}(z_t; \gamma, \mu)$ has the shape of a quadratic looking function in z_t . For very large values of γ , $\mathcal{G}(z_t; \gamma, \mu)$ is equal to 1 for all values of z_t , except for 3 points, of which 2 are close to 1, and 1 is exactly equal to 0.

III. Identification and estimation issues

We begin by illustrating that identification problems arise when $\mathcal{G}(z_t; \gamma, \mu)$ is well approximated by a quadratic function in z_t . Consider the general ESTAR specification for y_t in (1), where for simplicity and without loss of generality, we can restrict α to $\mathbf{0}$, to yield:³

$$y_t = x_t \beta \mathcal{G}(z_t; \gamma, \mu) + \epsilon_t, \tag{2a}$$

$$\mathcal{G}(z_t; \gamma, \mu) = 1 - \exp\{-\gamma(z_t - \mu)^2\}. \tag{2b}$$

applications, the appropriate regressor set for each regime is commonly determined by variable selection procedures (see van Dijk *et al.* (2002) for additional discussion) and is allowed to differ between the two regimes.

³Alternatively, one can restrict α at $\bar{\alpha}$, define $\tilde{y}_t = y_t - x_t \bar{\alpha}$, and then proceed as in (2) above, but now with \tilde{y}_t in place of y_t .

Expanding $\mathcal{G}(z_t; \gamma, \mu)$ around $\gamma = 0$ by a second-order Taylor series (denoted by $\tilde{\mathcal{G}}_2(z_t; \gamma, \mu)$), yields:

$$\tilde{\mathcal{G}}_2(z_t; \gamma, \mu) = \gamma(z_t - \mu)^2 - \frac{1}{2}\gamma^2(z_t - \mu)^4 + \mathcal{R}_2(z_t; \gamma, \mu), \tag{3}$$

and replacing $\mathcal{G}(z_t; \gamma, \mu)$ in (2) with $\tilde{\mathcal{G}}_2(z_t; \gamma, \mu)$ leads to the auxiliary model:

$$y_t = (z_t - \mu)^2 \mathbf{x}_t \boldsymbol{\beta} \gamma - \frac{1}{2}(z_t - \mu)^4 \mathbf{x}_t \boldsymbol{\beta} \gamma^2 + v_t, \tag{4}$$

where $v_t = \mathbf{x}_t \boldsymbol{\beta} \mathcal{R}_2(z_t; \gamma, \mu) + \epsilon_t$, and $\mathcal{R}_2(z_t; \gamma, \mu)$ is the remainder term taking the form $\sum_{j=3}^{\infty} (-1)^{j-1} \frac{1}{j!} \gamma^j (z_t - \mu)^{2j}$. The relation in (4) becomes a regression model with two sets of regression vectors, i.e., $(z_t - \mu)^2 \mathbf{x}_t$ and $(z_t - \mu)^4 \mathbf{x}_t$, and two sets of slope parameters, $\mathbf{a} = \boldsymbol{\beta} \gamma$ and $\mathbf{b} = -\frac{1}{2} \boldsymbol{\beta} \gamma^2$, which can be written as:

$$y_t = (z_t - \mu)^2 \mathbf{x}_t \mathbf{a} + \underbrace{(z_t - \mu)^4 \mathbf{x}_t \mathbf{b}}_{\text{higher order terms}} + v_t, \tag{5}$$

where we can notice that $\boldsymbol{\beta}$ and γ enter as a product into \mathbf{a} . Suppose that \mathbf{a} and \mathbf{b} are known. To be able to identify the separate effects of γ and $\boldsymbol{\beta}$ on the conditional mean we need to be able to recover γ and $\boldsymbol{\beta}$ from \mathbf{a} and \mathbf{b} . This is the model source of identification. We can see that $\mathbf{a} = \boldsymbol{\beta} \gamma$ and $\mathbf{b} = -\frac{1}{2} \boldsymbol{\beta} \gamma^2$ implies that $\mathbf{a} \gamma = -2\mathbf{b}$, which is a system of k equations so that γ can be recovered from either one of the relations $\gamma = -2 \frac{b_i}{a_i}, \forall i = 1, \dots, k$. Once γ is known, $\boldsymbol{\beta}$ can be obtained from $\boldsymbol{\beta} = \mathbf{a} / \gamma$. Thus, provided that at least one of the b_i coefficients on the higher order terms in (5) is non-zero, the model is identified. If this is not the case, the ESTAR model is not identified.

To formalise the concept of identification, let $\boldsymbol{\theta} = [\boldsymbol{\alpha}; \boldsymbol{\beta}; \gamma; \mu]$ be the parameter vector of interest with dimension $[(2k + 2) \times 1]$, and let \mathbf{Y} denote the full vector of observable random variables needed to formulate a probabilistic model for $\{y_t\}_{t=1}^T$, that is, $\mathbf{Y} = \{y_t, \mathbf{x}_t, z_t\}_{t=1}^T$. The likelihood function that describes the probabilistic model is then denoted by $p(\mathbf{Y}; \boldsymbol{\theta})$. Further, let Θ represent the admissible parameter space for $\boldsymbol{\theta}$, so that $\boldsymbol{\theta} \in \Theta$. Following from Definition 2 in Rothenberg (1971, page 578), a parameter $\boldsymbol{\theta}_1 \in \Theta$ is said to be globally identifiable if there exists no other $\boldsymbol{\theta}_2 \in \Theta$ which is observationally equivalent. More specifically, $\boldsymbol{\theta}_1 \in \Theta$ is said to be globally identifiable if for any other $\boldsymbol{\theta}_2 \in \Theta$, we have that $p(\mathbf{Y}; \boldsymbol{\theta}_1) \neq p(\mathbf{Y}; \boldsymbol{\theta}_2)$ for some observable data \mathbf{Y} . To obtain local identification, we only require $p(\mathbf{Y}; \boldsymbol{\theta}_1)$ to be unique around some neighbourhood of $\boldsymbol{\theta}_1$.

From Theorem 1 in Rothenberg (1971, page 579), it follows that a necessary and sufficient condition for local identifiability of $\boldsymbol{\theta}$ is that the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$ is non-singular when evaluated at the true $\boldsymbol{\theta}_0$ parameter value, where

$$\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E} \left[\left(\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta} | \mathbf{Y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \right], \tag{6}$$

$\mathbb{E}[\cdot]$ is the expectation operator, and $\mathcal{L}(\boldsymbol{\theta} | \mathbf{Y}) = \log[p(\mathbf{Y}; \boldsymbol{\theta})]$ is the log of the likelihood function. Local identification of the ESTAR model can then be determined by checking the rank of $\mathcal{I}(\boldsymbol{\theta})$ for all admissible points in the parameter space Θ .

Identification in models fitted to empirical data

In empirical data, identification is harder to determine exactly, as it will depend not only on the parameter combination, but also on the random variation in the sample. This means that $\mathcal{I}(\theta)$ can be nearly singular or rank deficient. Consequences of near singularity will be that small changes in the sample data \mathbf{Y} or the parameter vector θ can lead to large changes in the estimates of θ . Moreover, as is well known with weakly identified models, parameter estimates will be highly correlated, resulting in large standard errors, and finite sample distributions can be very different from their asymptotic approximations (see Pagan and Robertson (1998)).

To examine issues related to weak identification, we formulate a particular *true* parameter scenario θ_0 and then assess numerically how close to singularity $\mathcal{I}(\theta)$ is, conditional on the observed data \mathbf{Y} . This proximity to singularity can be measured by the size of the condition number of the information matrix $\mathcal{I}(\theta)$. For a non-singular matrix \mathcal{A} , the condition number is defined as:

$$\text{cond}(\mathcal{A}) = \|\mathcal{A}\| \|\mathcal{A}^{-1}\|, \quad (7)$$

where $\|\cdot\|$ denotes the norm. If the Euclidean or 2-norm is used, this becomes $\text{cond}_2(\mathcal{A}) = \lambda_{\max}/\lambda_{\min}$, where λ denotes the singular value of \mathcal{A} . When λ_{\min} is small and λ_{\max} is large, then $\text{cond}_2(\mathcal{A})$ can become extremely large, indicating that the inverse of matrix \mathcal{A} is badly conditioned. In general, when $\text{cond}_2(\mathcal{A})$ is large, then \mathcal{A} will be nearly singular, which in our case will mean weak identification. The closer $\text{cond}_2(\mathcal{A})$ is to its minimum value of 1, the further away \mathcal{A} is from singularity, and the stronger identified the model is.

How large a condition number is needed for \mathcal{A} to be considered singular? There exist not direct guidelines in the numerical computing literature to narrow down the magnitude of values, as it depends on the numerical precision of the computing environment. For instance, in Matlab (64 bit), the machine precision is double, that is, 2.2204×10^{-16} .⁴ In the econometrics literature, (Greene, 2011, p. 999) suggests that if $\text{cond}_2(\mathcal{A})$ is greater than $20^2 = 400$, then the condition number is considered to be extremely large and \mathcal{A} is singular numerically. In what follows below, we will use the condition number of the information matrix $\mathcal{I}(\theta)$ to determine how *strongly* identified the ESTAR model is for different values of the γ parameter and the observed data \mathbf{Y} .

Another quantity that is frequently computed to examine identification issues in the literature on DSGE models is the correlation matrix (see Iskrev, 2010). This matrix is defined as:

$$\tilde{\mathcal{I}}(\theta) = \mathcal{D}(\theta)^{-1/2} \mathcal{I}(\theta) \mathcal{D}(\theta)^{-1/2}, \quad (8)$$

where $\mathcal{D}(\theta) = \text{diag}[\mathcal{I}(\theta)]$. The correlation matrix is useful as it provides extra insights into which components of the parameter vector θ are causing identification problems. The condition number condenses the information down to one value to determine whether

⁴The function `eps` in Matlab provides this numerical value. Whenever the inverse or reciprocal of the 1-norm is less than machine precision, matrix \mathcal{A} is deemed to be poorly conditioned and a warning of the form: 'Warning: Matrix is close to singular or badly scaled. Results may be inaccurate' will be displayed. In Matlab, the LAPACK reciprocal condition number is called by the function `rcond`. Also, for a positive definite matrix \mathcal{A} , all the singular values will be positive and so the smallest value that $\text{cond}_2(\mathcal{A})$ can take is 1.

identification in a model could be a problem, but does not reveal which parameters are affected.

Spurious fitting of outliers

One other problem that frequently arises in the estimation of ESTAR models is the spurious overfitting of a small number of ‘*aberrant*’ observations around μ , which occurs when γ becomes extremely large. This can be of particular concern when working with empirical financial and/or macroeconomic time series data. The problem is that the exponential regime weighting function is equal to 1 in the limit when $\gamma \rightarrow \infty$ for all values of z_t , except at $z_t = \mu$. In finite samples it is nearly always possible to find a data point (or set of points) that needs to be downweighed to improve the fit of a linear model. In such instances, the exponential function acts like dummy variable, accommodating as good as possible the features of only a small number of observations around μ . We illustrate how this is omnipresent in the empirical real exchange rate data analysed in Taylor *et al.* (2001), but arises also when fitting ESTAR models to simulated data, which are well behaved and outlier free by construction.

IV. Empirical and simulation evidence

To provide evidence of the above outlined problems with ESTAR models, we begin with a replication of the well known and widely cited study by Taylor *et al.* (2001) on the speed of mean reversion in real exchange rates. We then proceed with a simulation analysis and examine how increasing the magnitude of γ and the sample size impact on the ‘*strength*’ of identification in ESTAR models, as well as the ‘*outlier fitting*’ problem. In order to keep this section as concise as possible, we have delegated a detailed description of the real exchange rate data used in Taylor *et al.* (2001) as well as additional estimation results for all 4 real exchange rate series to the Online Appendix of this paper.⁵

Replication of Taylor *et al.* (2001)

Taylor *et al.* (2001) fit the following ESTAR model to real exchange rate q_t :⁶

$$\Delta q_t = (\alpha - 1)(q_{t-1} - \mu) + \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t \tag{9a}$$

$$\mathcal{G}(q_{t-1}; \gamma, \mu) = 1 - \exp \{ -\gamma(q_{t-1} - \mu)^2 \}, \tag{9b}$$

where ϵ_t is a zero mean disturbance term with variance σ^2 . Through the use of standard statistical tests, Taylor *et al.* (2001, page 1030) conclude that: “... *in no case could we reject at the five percent significance level the restrictions that $\alpha - \beta = 1$.*” The final restricted model that is estimated is:

⁵ In section A.5. of the Online Appendix, we repeat the analysis presented here and illustrate how the same problems also arise in another study by Teräsvirta and Anderson (1992) who estimate a ESTAR models for Japanese and Italian industrial production data.

⁶ We have re-written the model by subtracting $(q_{t-1} - \mu)$ from both sides so that the results can be interpreted and visualised with Δq_t as the response variable on the left hand side, and when plotted on the y - axis.

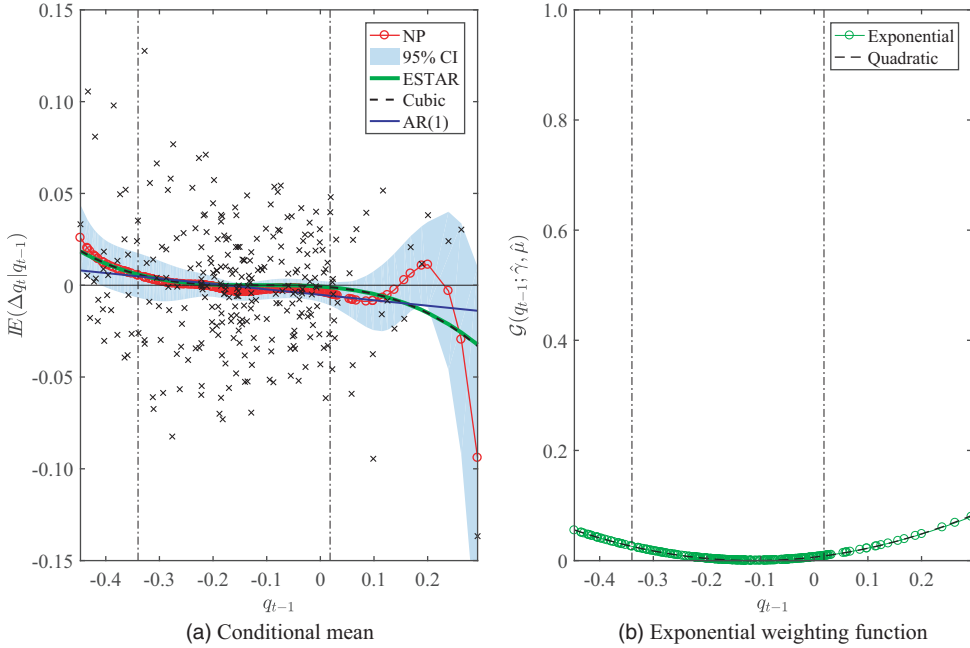


Figure 2. Conditional mean and weighting function of the restricted ESTAR model for the UK series

$$\Delta q_t = -(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t. \tag{10}$$

Full replication results for all 4 currency pairs are reported in Table A.1 in the Online Appendix. The parameter estimates that we obtain are nearly identical to those listed in Table 3 on page 1029 in Taylor *et al.* (2001), with the estimates of γ being in the 0.183 to 0.505 range.

To show visually what these γ estimates imply for the conditional mean of the model and the regime weighting function, we plot the conditional mean and weighting function for the UK real exchange series in Figure 2.⁷ The green line (with circles) shows the implied conditional mean and weighting function at the ESTAR parameter estimates, while the black dashed lines show corresponding cubic and quadratic fits in $(q_{t-1} - \mu)$. Examining the plot of the weighting function in Panel (b) of Figure 2, we can see how flat and weakly curved the weighting function is over the range of the threshold variable. Such a shape can be well approximated by a quadratic function in $(q_{t-1} - \mu)$. The fitted model is thus likely to be unidentified.

To formally examine how *strongly* identified the model is with respect to the γ and β parameters at their estimates, we compute the condition number of the information

⁷We use the results for the UK series here as a representative example to illustrate our point, and report results for all 4 series in the Online Appendix. The γ estimates for the UK series is the largest, so in a sense represents the most nonlinear ESTAR model of the 4 series. In the conditional mean plot in Panel (a), we also superimpose a linear AR(1) fit, a non-parametric fit using a local linear kernel regression estimate with 95% confidence bands, as well as a scatter of the data.

TABLE 1
Correlation matrix $\tilde{\mathcal{I}}_3(\hat{\theta})$

UK	Germany	France	Japan
$\begin{bmatrix} 1 & - & - \\ -0.9948 & 1 & - \\ -0.2346 & -0.2371 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & - & - \\ -0.9949 & 1 & - \\ -0.4365 & 0.4412 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & - & - \\ -0.9952 & 1 & - \\ -0.4132 & -0.4171 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & - & - \\ -0.9959 & 1 & - \\ -0.4010 & -0.4053 & 1 \end{bmatrix}$

Notes: This table reports the correlation matrix $\tilde{\mathcal{I}}_3(\hat{\theta})$ of the parameter estimates with $\bar{\beta} = -1$ and γ and μ evaluated at their ML estimates. The parameter ordering of the $\tilde{\mathcal{I}}_3(\hat{\theta})$ matrix is $[\bar{\beta}; \hat{\gamma}; \hat{\mu}]$.

matrix as outlined in section III. To illustrate this computation, let $\mathcal{I}_3(\hat{\theta})$ denote the (3×3) dimensional information matrix, where $\hat{\theta} = [\bar{\beta}; \hat{\gamma}; \hat{\mu}]$, with $\bar{\beta} = -1$, and $\hat{\gamma}$ and $\hat{\mu}$ at their estimates. Computing the condition number of $\mathcal{I}_3(\hat{\theta})$ as the ratio of the largest to the smallest singular values for the four countries of interest yields values of 833, 1,436, 1,507, and 1,875 for the UK, Germany, France and Japan, respectively.⁸ These condition numbers are substantially larger than the threshold value of $20^2 = 400$ suggested in Greene (2011).

The correlation matrix of the information matrix at $\hat{\theta} = [\bar{\beta}; \hat{\gamma}; \hat{\mu}]$, is shown in Table 1 below. The correlation between the first two elements of $\hat{\theta}$ corresponding to the β and γ parameters exceeds 0.99 in absolute value for all 4 real exchange rate series. Movements in these two conditional mean parameters off-set one another nearly one for one. These results thus confirm our previous claim that the estimated ESTAR model is not identified.

Can we estimate the model without the $\beta = -1$ restriction?

We predict that $\hat{\gamma} \rightarrow 0$, and $\hat{\beta}$ becomes very large in absolute value to accommodate the cubic structure of the ESTAR conditional mean of the model. Estimation results for the following unrestricted model:

$$\Delta q_t = \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t \tag{11}$$

are reported in Table 2. As can be seen from the β and γ estimates, and as predicted, there is an off-setting effect. The estimate of γ converges towards 0, while $\hat{\beta}$ goes towards a large negative value. Note that the lower bound in the initial grid search for $\hat{\gamma}$ was set at 1×10^{-6} , so the values that are estimated are only marginally lower than that. What is important to point out here is that the log-likelihood function of the ESTAR and the cubic models are numerically identical up to 6 decimal points. Relating this to the identification discussion of section 3, the situation that we have here is $\hat{\gamma} \rightarrow 0$, $\hat{\beta} \rightarrow -\infty$, with $\hat{\beta}\hat{\gamma} \rightarrow \hat{\delta}_3 \neq 0$, where $\hat{\delta}_3$ is the coefficient obtained from a cubic regression, ie., of Δq_t on $(q_{t-1} - \mu)^3$. The higher order terms in the second-order approximation listed in (4) related to $\hat{\beta}\hat{\gamma}^2$ go to 0 due to $\hat{\gamma} \rightarrow 0$. Given the information in the data, the conditional means of the cubic regression

⁸These condition numbers were computed after re-scaling the columns of $\mathcal{I}_3(\hat{\theta})$ by each column's norm to have unit length. This is done to avoid any issues related to the scaling of the information matrix (see also the discussion on page 999 in Greene (2011)).

TABLE 2
Parameter estimates of the ESTAR model without the $\beta = -1$ restriction

Parameter Estimates	UK	Germany	France	Japan
$\hat{\beta}$	-504662.45687917073000	-305284.18535727740000	-368394.40126472880000	-184561.06430017314000
$\hat{\gamma}$	0.00000098499515	0.00000095653641	0.00000095237141	0.00000098340592
$\hat{\mu}$	-0.11230079305474	0.00924026196683	-0.00571998249191	-0.51121351057606
$\hat{\sigma}$	0.03319912785833	0.03448623763652	0.03286238780413	0.03330711916822
Log-like. ESTAR	570.06613104441669	559.14958978825916	572.99204734048169	569.13408226512229
Log-like. Cubic model	570.06613110866749	559.14958988203637	572.99204741222627	569.13408243358822
Log-like. AR(1) model	568.99987094204778	557.92278404913804	571.80528152106854	567.46500060788514

Notes: This table reports the replicated parameter estimates of the ESTAR model of Taylor *et al.* (2001) which only imposes the unit-root restriction on the inner regime. This ESTAR model takes the form: $\Delta q_t = \beta(q_{t-1} - \mu) \mathcal{G}(q_{t-1}; \gamma; \mu) + \epsilon_t$. Log-likelihood functions of the AR(1) and Cubic models are based on the two regression equations: $\Delta q_t = \delta_1(q_{t-1} - \mu) + \epsilon_t$ and $\Delta q_t = \delta_3(q_{t-1} - \mu)^3 + \epsilon_t$.

TABLE 3
Parameter estimates of the unrestricted ESTAR model

Parameter Estimates	UK	Germany	France	Japan
$(\hat{\alpha} - 1)$	14183.23931548780800	-80.43652136645221	4895.09914479176040	340594.58805191959000
$\hat{\beta}$	-14183.25851543536600	80.41534978248009	-4895.11461596329630	-340594.59946709510000
$\hat{\gamma}$	1937108.34779812440000	220354.95060880922000	1129966.10535951520000	1587721.65109557380000
$\hat{\mu}$	-0.27437230860580	-0.15258839513024	-0.02174838959589	-0.70815477313544
$\hat{\sigma}$	0.03272229911343	0.03387140750452	0.03223190772543	0.03256410350725
Log-like. ESTAR	574.21811144990806	564.31246531012380	578.55178002589184	575.60897466208894
Log-like. Cubic model	570.06613110866749	559.14958988203637	572.99204741222627	569.13408243358822
Log-like. AR(1) model	568.99987094204778	557.92278404913816	571.80528152106854	567.46500060788514

Notes: This table reports the replicated parameter estimates of the unrestricted ESTAR model of Taylor *et al.* (2001). The unrestricted ESTAR model takes the form: $\Delta q_t = (\alpha - 1)(q_{t-1} - \mu) + \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t$. Log-likelihood functions of the AR(1) and Cubic models are as reported before.

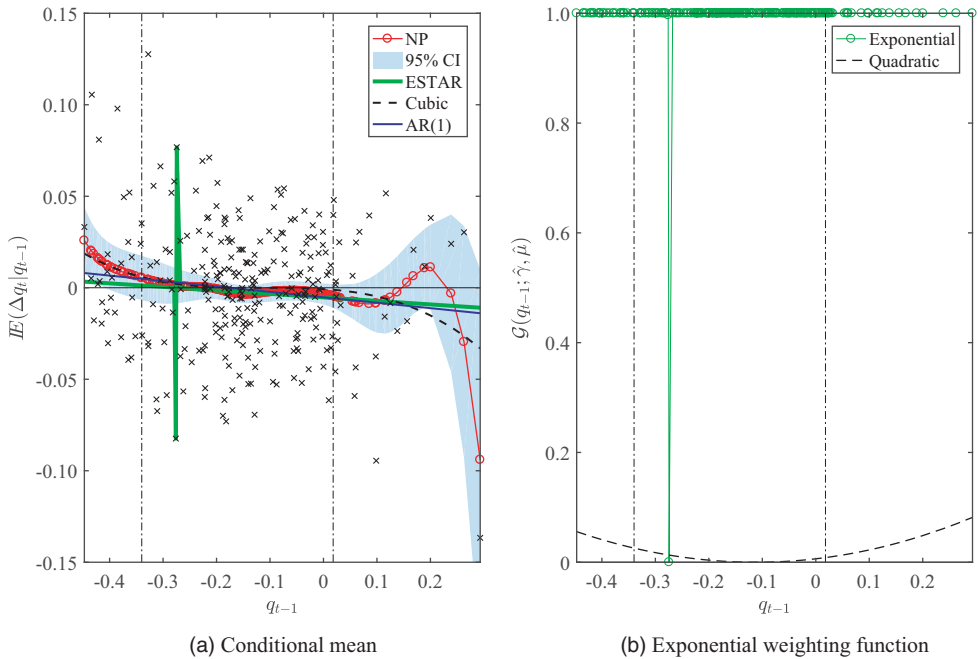


Figure 3. Conditional mean and weighting function of the unrestricted ESTAR model for the UK series

and the ESTAR model provide the same fit. This is also evident visually from Panel (a) of Figure 2.

Estimating the unrestricted model

To illustrate that the exponential weighting function has a propensity to overfit outliers by acting like a dummy variable function, we estimate the unrestricted ESTAR model of Taylor *et al.* (2001) defined in (9a). These estimation results are reported in Table 3 below.

As can be seen from the estimates, $\hat{\gamma}$ attains extremely large values. The slope coefficients $(\hat{\alpha} - 1)$ and $\hat{\beta}$ are mirror images of one another, in the sense that their absolute magnitudes are very similar, differing only in their sign. Notice here also that all log-likelihoods are larger than the estimates from the restricted models of Taylor *et al.* (2001).

To understand what these models try to capture in the data, we again plot the implied conditional mean as well as weighting function at the estimates for the UK real exchange rate series in Figure 3. The fitted weighting function in Panel (b) of Figure 3 is numerically equal to 1 (up to 8 decimal places) for all but two observations. One of these is 0.99668679, which is also quite close to 1, with the second being 0.00005427. The function thus assigns a zero weight to one observation only, and thus effectively acts as a dummy variable for that observation. The effect of this on the conditional mean is visible from Panel (a) of Figure 3. The conditional mean is a linear function of q_{t-1} over the entire range of the threshold variable, with the only exception being around $\hat{\mu}$, where it spikes up and down. The AR(1) conditional mean shown by the solid blue line in Panel (a) largely overlaps with the ESTAR model's fit.

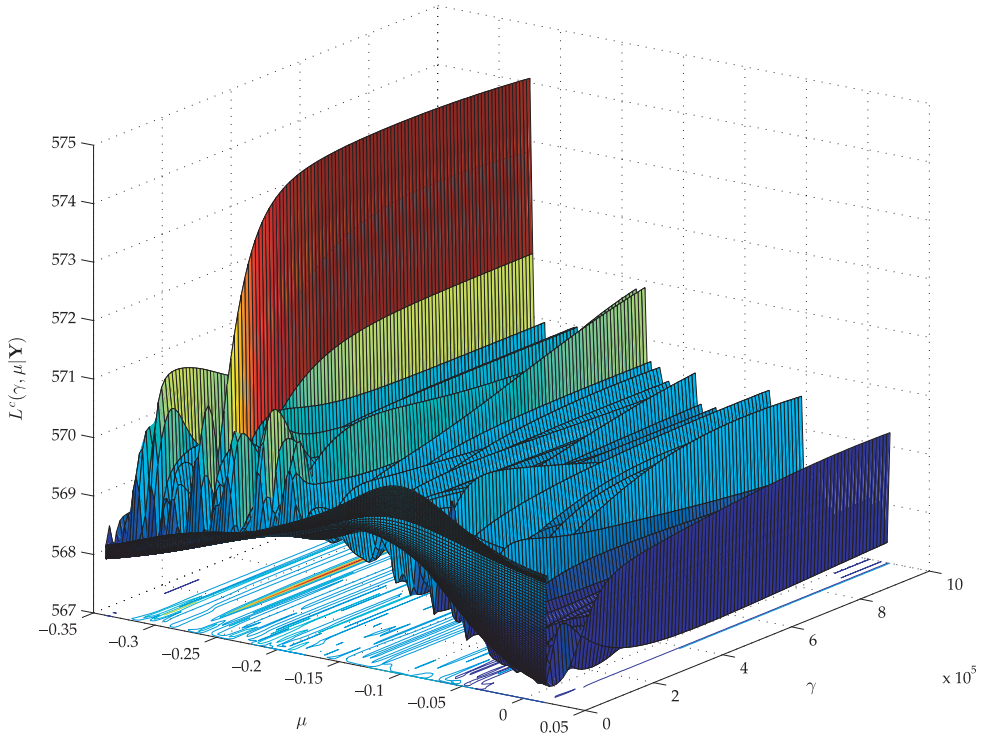


Figure 4. Concentrated log-likelihood function of the unrestricted ESTAR model for the UK series

In Figure 4, we show surface plots of the concentrated log-likelihood function over the γ and μ grids for the UK real exchange rate series. What is evident from these plots is that the concentrated log-likelihood function is extremely ill-behaved, with a large number of local maxima. This ill-behaviour is not only visible for extremely large values of γ , but also for more moderately sized values. Searching over the threshold location parameter μ allows for large γ values and therefore admits the fitting of a very few extreme observations, which results in the highly irregular shape of the concentrated log-likelihood function and the conditional means that we observe.

Simulation evidence

We proceed the analysis by examining the above outlined problems within a controlled simulation experiment. To keep the considered calibrations limited, we take the UK parameter estimates as our baseline data-generating process, and then consider increasing values of γ as additional parameter scenarios to examine its effect on identification. The model that we simulate from is given in (9). The true set of parameters used in the simulation are: $\mu_0 = -0.1124$, $\alpha_0 = 1$, $\beta_0 = -1$, and $\sigma_0 = 0.0332$. The error term is drawn from a standard Normal distribution. The six different γ values that we consider are $\gamma_0 = \{0.505, 5, 50, 250, 500, 1000\}$. The 0.505 value is the empirical estimate in the data, while the other values approximate various nonlinear shapes in the transition function and

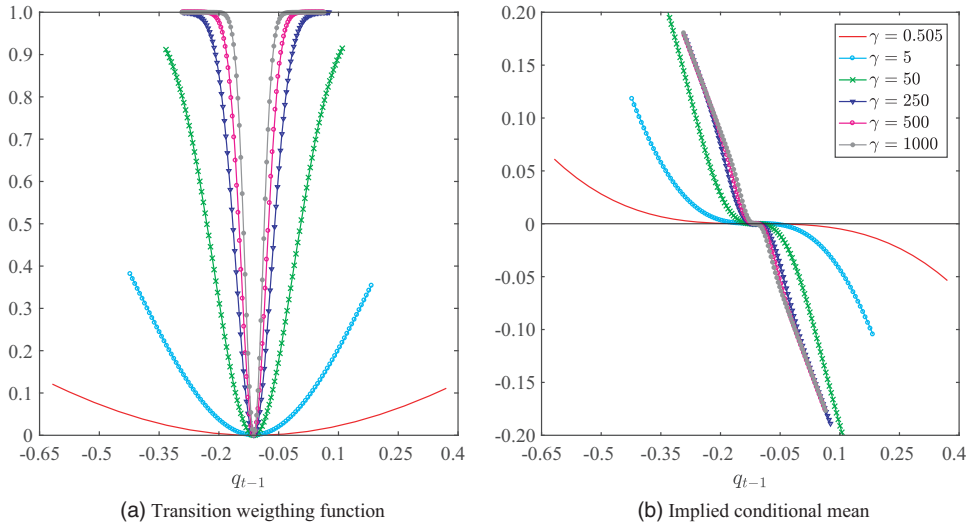


Figure 5. Exponential transition function weights $\mathcal{G}(z_t; \gamma, \mu)$ and implied conditional means of the ESTAR models that we simulate from. These are evaluated and plotted over an equally spaced grid of 101 points from $\min(q_t)$ to $\max(q_t)$ for each of the considered γ_0 grid values

also the implied conditional mean of the ESTAR model. In total, we simulate $S = 10,000$ sequences, with samples of size $T = \{288, 500, 1, 500, 5, 000\}$.

As the shape of the transition weighting function is important for understanding identification issues in ESTAR models, we provide here a visual indication of the various shapes that $\mathcal{G}(q_{t-1}; \gamma, \mu)$ and the conditional mean can take under the above considered true model calibrations. These plots are shown in the left and right hand panels of Figure 5, respectively. Each of the weighting function and conditional mean plots under the different γ_0 values are drawn over the maximal range of the simulated q_t series.

From the plots in Figure 5, we can see that the lowest γ value generates the widest range for q_{t-1} , while for larger γ values, the q_{t-1} range becomes narrower. This is particularly evident for γ values of 250, 500 and 1,000, where the transition weighting function covers all feasible values in the $[0, 1]$ interval of $\mathcal{G}(q_{t-1}; \gamma, \mu)$, with the q_{t-1} range being rather narrow between -0.35 and 0.1 . The key point to take away from these plots is that the calibrations that we choose for the simulations cover a wide range of possible shapes of the transition weighting function $\mathcal{G}(q_{t-1}; \gamma, \mu)$. As the shape of $\mathcal{G}(q_{t-1}; \gamma, \mu)$ determines how weakly identified the ESTAR model is, we can see that there is a wide coverage of different shapes and hence identification scenarios.

Identification results

In Figure 5, we report arithmetic averages of the correlations between the β and γ parameters and the condition numbers of the (3×3) information matrix $\mathcal{I}_3(\hat{\theta})$, where we follow again the same computational approach as in Section 4.1 with the empirical data. The estimated ESTAR model is the restricted model in (10). Row entries in Table 4 correspond to the six different γ_0 calibrations that are used, i.e., $\{0.505, 5, 50, 250, 500, 1, 000\}$, while column entries list the considered sample sizes $T = \{288, 500, 1, 500, 5, 000\}$.

TABLE 4
Correlation and condition numbers

$\gamma_0 \backslash T$	288		500		1500		5,000	
	$corr(\beta, \gamma)$	$cond_2(\mathcal{I}_3)$	$corr(\beta, \gamma)$	$cond_2(\mathcal{I}_3)$	$corr(\beta, \gamma)$	$cond_2(\mathcal{I}_3)$	$corr(\beta, \gamma)$	$cond_2(\mathcal{I}_3)$
0.505	-0.9970	7,241.1915	-0.9978	4,780.6845	-0.9987	10,686.7550	-0.9993	14,630.1525
5	-0.9939	1,607.4277	-0.9953	2,201.8775	-0.9971	7,583.2263	-0.9983	5,614.7110
50	-0.9780	606.5410	-0.9818	844.8802	-0.9867	520.5987	-0.9889	349.0564
250	-0.9011	57.3385	-0.9065	40.7965	-0.9086	23.2062	-0.9089	21.5003
500	-0.8064	15.1008	-0.8108	13.0014	-0.8125	10.1182	-0.8127	9.7991
1,000	-0.6634	6.6035	-0.6679	5.6554	-0.6711	5.2339	-0.6717	5.1359

Notes: This table reports averages of the correlation between the β and γ parameter estimates and the condition number of the (3×3) information matrix $\mathcal{I}_3(\hat{\theta})$ for various sample sizes and γ_0 values. We consider γ_0 values over the grid $\in \{0.505, 5, 50, 250, 500, 1,000\}$ and sample sizes of $T \in \{288, 500, 1,500, 5,000\}$. These results are based on arithmetic averages computed over 10,000 simulations. The correlation between β and γ , denoted by $corr(\beta, \gamma)$, is computed as the $\{2, 1\}$ element of the correlation matrix $\tilde{\mathcal{I}}_3(\hat{\theta})$ defined in (8), with $\hat{\theta} = [\hat{\beta}; \hat{\gamma}; \hat{\mu}]$ and $\hat{\beta} = -1$. The condition number, denoted by $cond_2(\mathcal{I}_3)$ in the table, is also evaluated at $\hat{\theta}$.

From the correlation and condition number entries in Table 4, we can see that for the smaller γ_0 values, the identification problem is not a finite sample issue, but persists for all sample sizes we consider. For instance, when $\gamma_0 = 0.505$ as in Taylor *et al.* (2001), the average correlation between $\hat{\beta}$ and $\hat{\gamma}$ increases (in absolute value) steadily toward 1 as the sample size increases from 288 observations to 5,000. At the same time, the condition number more than doubles, suggesting that the model becomes ‘less identified’ as the sample size increases. As γ_0 increases in size, the correlations as well as the condition numbers decrease steadily. This result is most clearly seen from the change in the size of the correlations and condition numbers when moving from a γ_0 value of 50 to 250. The correlations decrease in absolute value from around 0.98 to about 0.90, with the condition numbers dropping more than tenfold to values of about 57 or less, which are well below the threshold of $20^2 = 400$ suggested by Greene (2011). For γ_0 values of 5 and 50, the correlations and condition numbers are still rather large, while for higher γ_0 values of 500 and 1,000, these numbers decrease, suggesting that β and γ can now be identified from the data.

Overall, the results presented here confirm our hypothesis of identification problems in ESTAR models being linked to the magnitude of the γ parameter and the shape of the transition weighting function $\mathcal{G}(q_{t-1}; \gamma, \mu)$.

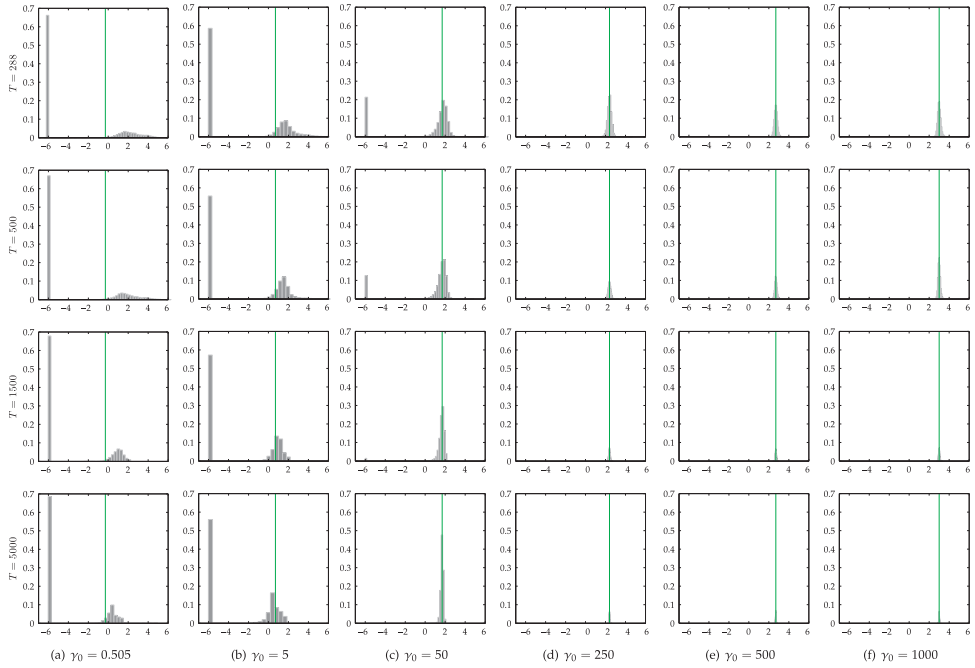
Estimating the unrestricted models

We now illustrate what happens when attempting to estimate the two unrestricted models of Taylor *et al.* (2001), that is,

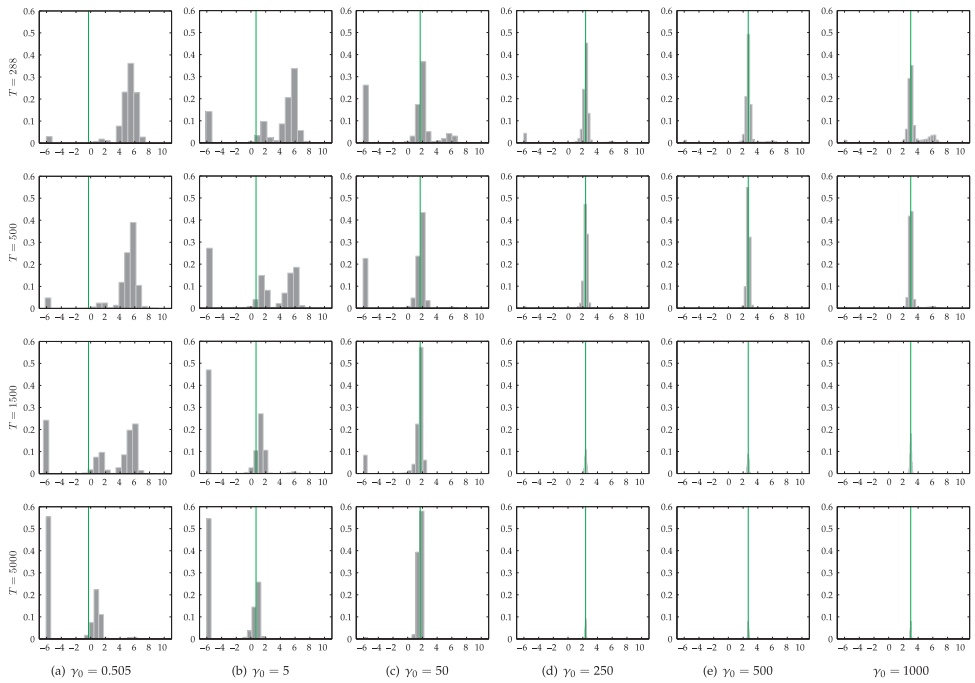
$$\Delta q_t = \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t, \tag{12}$$

$$\Delta q_t = (\alpha - 1)(q_{t-1} - \mu) + \beta(q_{t-1} - \mu)\mathcal{G}(q_{t-1}; \gamma, \mu) + \epsilon_t, \tag{13}$$

on the simulated data. Since the transition function parameter is key in determining the shape of $\mathcal{G}(q_{t-1}; \gamma, \mu)$, and therefore the stability as well as identification in the ESTAR model, we report estimation results for γ only.



(a) ESTAR model without the $\beta = -1$ restriction in (12)



(b) Unrestricted ESTAR model in (13)

Figure 6. Relative frequency plots of \log_{10} transformed γ estimates for various sample sizes and γ_0 values. Vertical green lines mark the true γ_0 value. The x -axis is log to base 10 transformed, with a reading of -6 on the x -axis meaning 1×10^{-6}

In Figure 6, we show relative frequency plots of the log to base 10 transformed γ estimates from fitting the two unrestricted models defined in (12) and (13) above, respectively, in Panels (a) and (b).⁹ The figures are arranged in 4 rows and 6 columns, each corresponding to the 4 different sample sizes and 6 γ_0 values that we simulate from. We superimpose a thin green vertical line to mark the true γ_0 value. Panel (a) of Figure 6 shows that for about 70% of the simulated ESTAR processes the estimates of the γ parameter hit the lower bound of the grid of 1×10^{-6} for all sample sizes when $\gamma_0 = 0.505$. With $\gamma_0 = 5$, this drops to about 57%. For $\gamma_0 = 50$, between 20% and 13% of the simulations hit the lower parameter bound for sample sizes of 288 and 500 observations, while for larger sample sizes, the estimates are concentrated closer around the true parameter value (in \log_{10} scale).

Panel (b) of Figure 6 shows analogous plots of $\hat{\gamma}$ estimated from the unrestricted ESTAR model in (13). At $\gamma_0 = 0.505$, the majority of estimates are in the 1×10^4 to 1×10^7 range, even for samples as large as 1,500. At $T = 5,000$, over 55% of the γ estimates hit the lower bound of 1×10^{-6} . When the true parameter is equal to 5, there remain a large number of extreme values that are obtained for $\hat{\gamma}$, that is, either extremely low or extremely high, with $\hat{\gamma}$ converging to 0 about 50% of the time for samples of size 1,500 and 5,000. As the sample size and true γ_0 values increase, the estimates tend to stabilize in the sense that they are away from the bounds and centre at γ_0 . All in all, the simulation results mirror our earlier findings based on the empirical real exchange rate data, that it is extremely difficult to estimate unrestricted ESTAR models due to the problematic shape the exponential function can take for small and large γ values.

One final point that we would like to make here is that the concentrated log-likelihood function of the unrestricted ESTAR model in (13) remains extremely ill-behaved with many local maxima and abrupt changes, even when the model is estimated on simulated data. In Figure A.7 in the Online Appendix, we show graphically the evolution of the concentrated log-likelihood surface for different γ_0 values and samples sizes to provide some visual evidence of this finding. The smoothness of the log-likelihood function tends to increase not only with the sample size, but also with the magnitude of the γ_0 value which generates the data.

V. Conclusion and suggestions for empirical work

We show that the exponential function is not suitable as a regime weighting function because of two undesirable properties. First, the exponential function can be well approximated by a quadratic function in the threshold variable z_t , whenever the transition function parameter γ takes on 'small' values. The consequence of this is that the slope vector attached to the nonlinear regime and the transition function parameter will enter as a product into the conditional mean of the model, which leads to identification issues. Using an empirical example and an extensive simulation analysis, we show that there is a near perfect off-setting effect of these two parameters. What is particularly problematic is that it does not vanish as the sample size increases.

⁹We plot \log_{10} transformed histograms of $\hat{\gamma}$ to be able to better illustrate graphically the large dispersion as well as the clustering at the boundaries of the estimates, and how these vanish with increasing γ_0 values and sample sizes. We have therefore intentionally left the axis scaling the same across the histogram plots.

Second, the exponential function can behave like an ‘outlier fitting function’. That is, for large values of γ , it will be equal to one for all values of the transition variable, except at $z_t = \mu$. When this occurs, only a very small number of observations around μ receive a weight different from one. The exponential function can thus act like a dummy variable which removes or downweights the influence of aberrant observations. Our empirical application shows that this is precisely the case with the study by Taylor *et al.* (2001). Estimating an unrestricted ESTAR model always returns an extremely large γ estimate, rendering the conditional mean a linear function of z_t for nearly its entire range. Using simulated data, we show that this occurs well over 70% of the time for the two smaller γ_0 values, and for sample sizes as large as 500 observations. Contrary to the identification issue, our results indicate that this problem vanishes as the sample size increases.

What alternative weighting functions exist that avoid the above outlined problems with the exponential function that an applied researcher or practitioner can use? We provide the interested reader with two simple existing alternatives that can be used in Section A.6. of the Online Appendix.

Final Manuscript Received: August 2018

References

- Baum, C. F., Barkoulas, J. T. and Caglayan, M. (2001). ‘Nonlinear adjustment to purchasing power parity in the post-bretton woods era’, *Journal of International Money and Finance*, Vol. 20, pp. 379–399.
- Beckmann, J., Berger, T. and Czudaj, R. (2015). ‘Does gold act as a hedge or a safe haven for stocks? A smooth transition approach’, *Economic Modelling*, Vol. 48, pp. 16–24.
- Buncic, D. (2012). ‘Understanding forecast failure of ESTAR models of real exchange rates’, *Empirical Economics*, Vol. 34, pp. 399–426.
- Buncic, D. and Piras, G. D. (2016). ‘Heterogenous agents, the financial crisis and exchange rate predictability’, *Journal of International Money and Finance*, Vol. 60, pp. 313–359.
- Cerrato, M., Kim, H. and Macdonald, R. (2010). ‘Three-regime asymmetric STAR modeling and exchange rate reversion’, *Journal of Money, Credit and Banking*, Vol. 42, pp. 1447–1467.
- Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- Greene, W. H. (2011). *Econometric Analysis*, 7th edn, Prentice Hall, Upper Saddle River, NJ.
- Haggan, V. and Ozaki, T. (1981). ‘Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model’, *Biometrika*, Vol. 68, pp. 189–196.
- Iskrev, N. (2010). ‘Local identification in DSGE models’, *Journal of Monetary Economics*, Vol. 57, pp. 189–202.
- Kapetanios, G., Shin, Y. and Snell, A. (2003). ‘Testing for a unit root in the nonlinear STAR framework’, *Journal of Econometrics*, Vol. 112, pp. 359–379.
- Kilian, L. and Taylor, M. P. (2003). ‘Why is it so difficult to beat the random walk forecast of exchange rates?’, *Journal of International Economics*, Vol. 60, pp. 85–107.
- Michael, P., Nobay, A. R. and Peel, D. A. (1997). ‘Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation’, *Journal of Political Economy*, Vol. 105, pp. 862–879.
- Pagan, A. R. and Robertson, J. C. (1998). ‘Structural models of the liquidity effect’, *The Review of Economics and Statistics*, Vol. 80, pp. 202–217.
- Pavlidis, E. G., Paya, I. and Peel, D. A. (2011). ‘Real exchange rates and time-varying trade costs’, *Journal of International Money and Finance*, Vol. 30, pp. 1157–1179.
- Paya, I. and Peel, D. A. (2006). ‘A new analysis of the determinants of the real Dollar-Sterling exchange rate: 1871–1994’, *Journal of Money, Credit and Banking*, Vol. 38, pp. 1971–1990.
- Rothenberg, T. J. (1971). ‘Identification in parametric models’, *Econometrica*, Vol. 39, pp. 577–591.

- Sarantis, N. (1999). 'Modeling non-linearities in real effective exchange rates', *Journal of International Money and Finance*, Vol. 18, pp. 27–45.
- Sarno, L., Leon, H. and Valente, G. (2006). 'Nonlinearity in deviations from uncovered interest parity: An explanation of the forward bias puzzle', *Review of Finance*, Vol. 10, pp. 321–482.
- Sollis, R. (2008). 'U.S. Dollar real exchange rates: Nonlinearity revisited', *Journal of International Money and Finance*, Vol. 27, pp. 516–528.
- Taylor, M. P. and Kim, H. (2009). 'Real variables, nonlinearity, and European real exchange rates', in Frankel, J. and Pissarides, C. (eds.), *NBER International Seminar on Macroeconomics 2008*, Chicago, IL: University of Chicago Press, pp. 157–193.
- Taylor, M. P. and Peel, D. A. (2000). 'Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals', *Journal of International Money and Finance*, Vol. 19, pp. 33–53.
- Taylor, M. P., Peel, D. A. and Sarno, L. (2001). 'Nonlinear mean-reversion in real exchange rates: Towards a solution to the purchasing power parity puzzles', *International Economic Review*, Vol. 42, pp. 1015–1042.
- Teräsvirta, T. (1994). 'Specification, estimation, and evaluation of smooth transition autoregressive models', *Journal of the American Statistical Association*, Vol. 89, pp. 208–218.
- Teräsvirta, T. and Anderson, H. M. (1992). 'Characterizing nonlinearities in business cycles using smooth transition autoregressive models', *Journal of Applied Econometrics*, Vol. 7, pp. 119–136.
- Tong, H. (1983). *Threshold models in non-linear time series analysis*, Springer Verlag, New York.
- van Dijk, D., Teräsvirta, T. and Franses, P. H. (2002). 'Smooth transition autoregressive models – A survey of recent developments', *Econometric Reviews*, Vol. 21, pp. 1–47.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Online appendix to identification and estimation issues in ESTAR models.